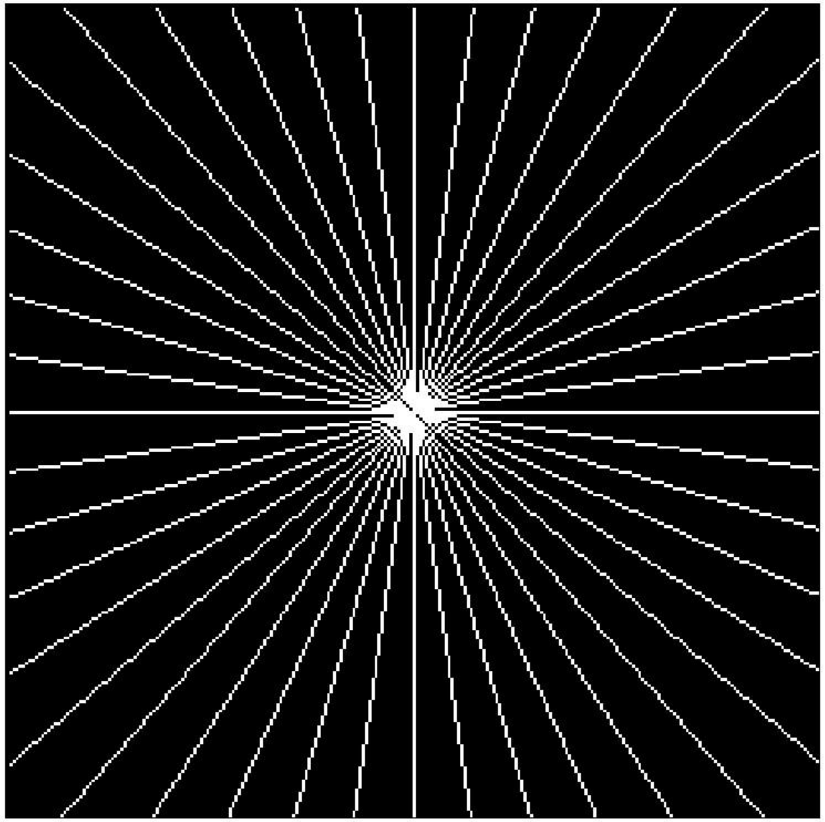
Compressive Sampling

Emmanuel Candès, California Institute of Technology

Colloquium Series, Los Alamos National Laboratories, October 2005

Collaborators: Justin Romberg (Caltech), Terence Tao (UCLA)

MRI Angiography Problem (Sketch)



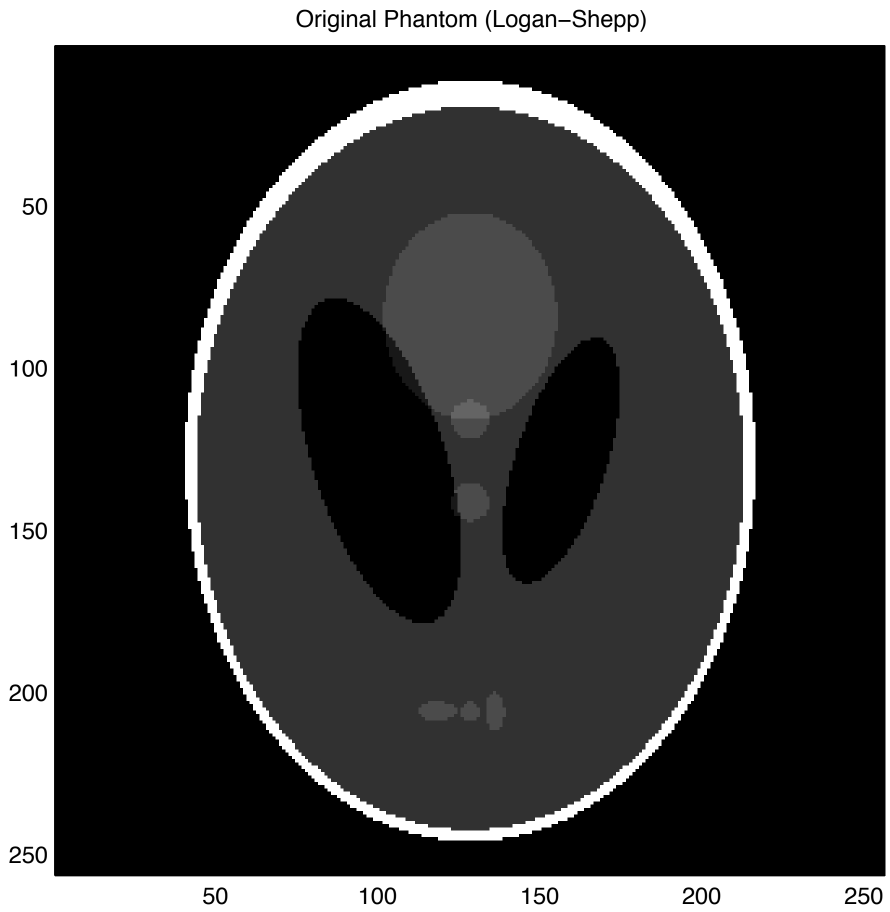
$\mathbf{22}$ radial lines

- $\approx \mathbf{4\%}$ coverage for 512 by 512 image (96% missing)
- $\approx \mathbf{2\%}$ coverage for 1024 by 1024 image (98% missing)

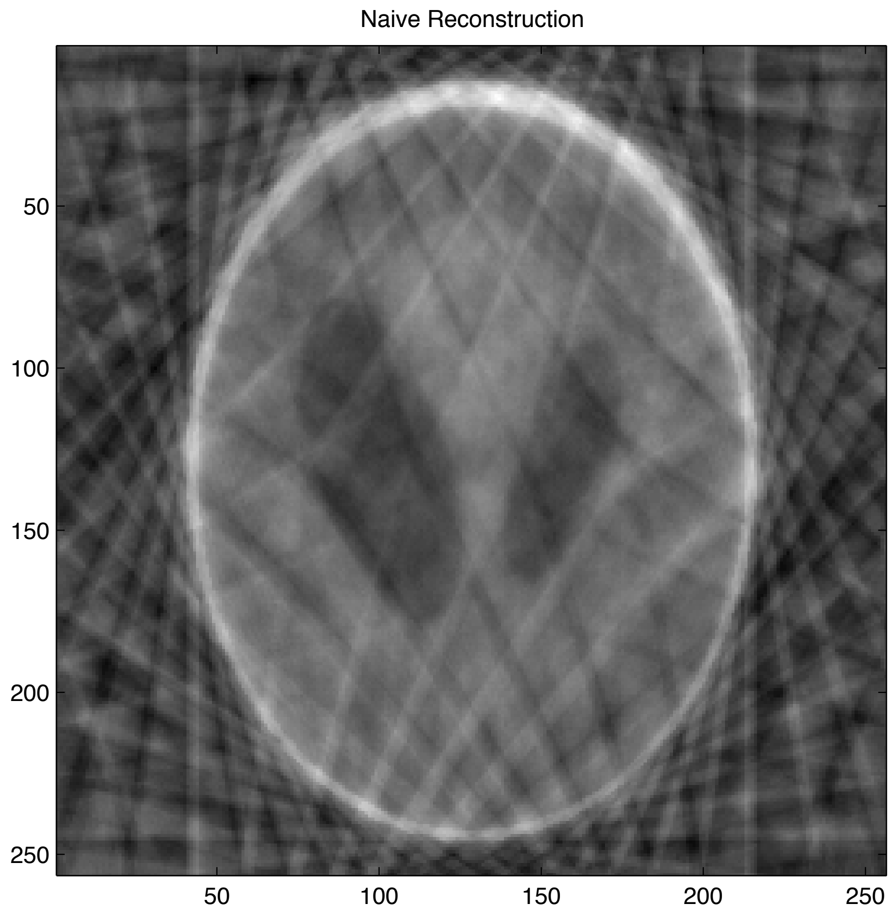
Classical Reconstruction

Backprojection: essentially reconstruct g^* with

$$\hat{g}^*(\omega) = \begin{cases} \hat{f}(\omega) & \omega \in \Omega \\ 0 & \omega \notin \Omega \end{cases}$$

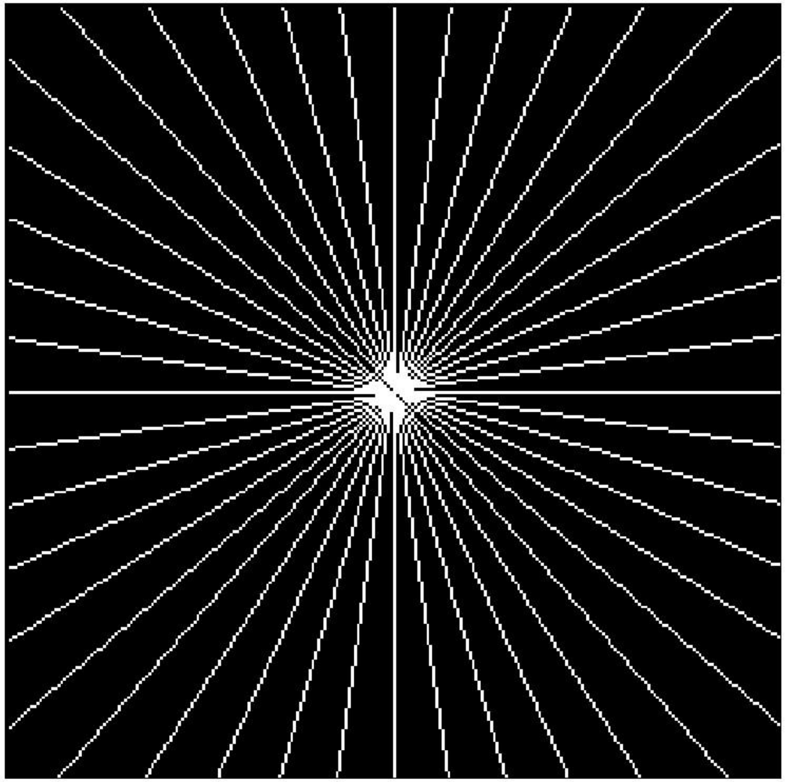


original

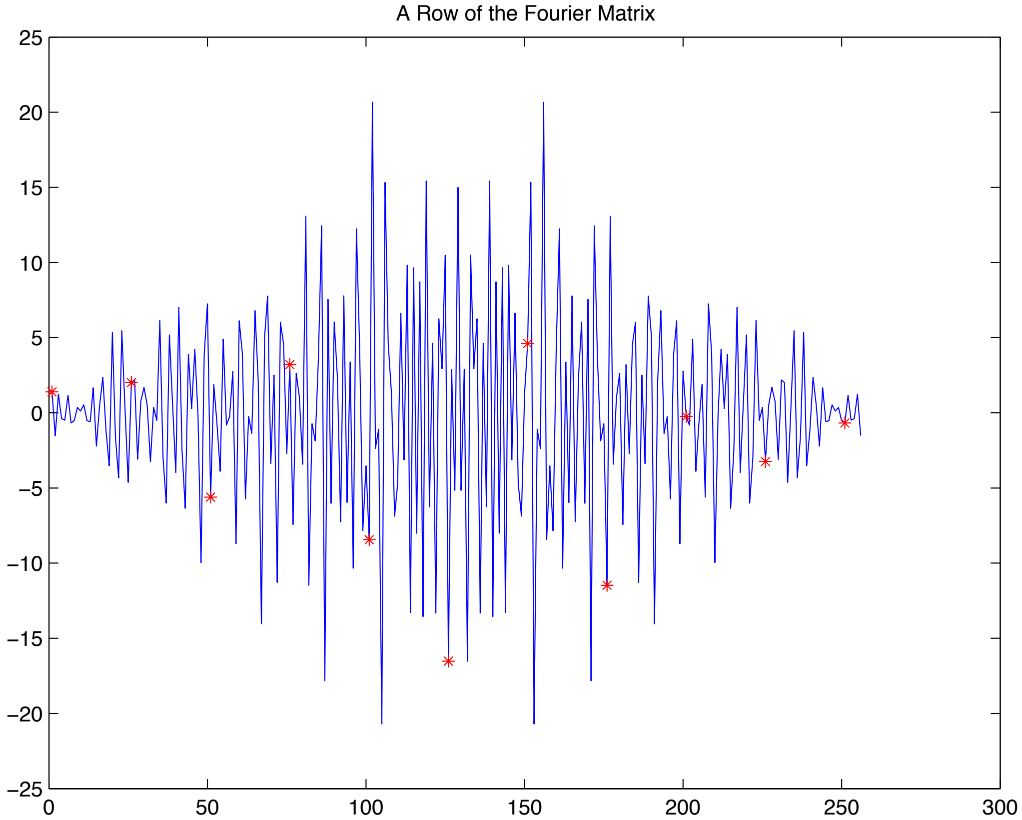


g^*

Interpolation?

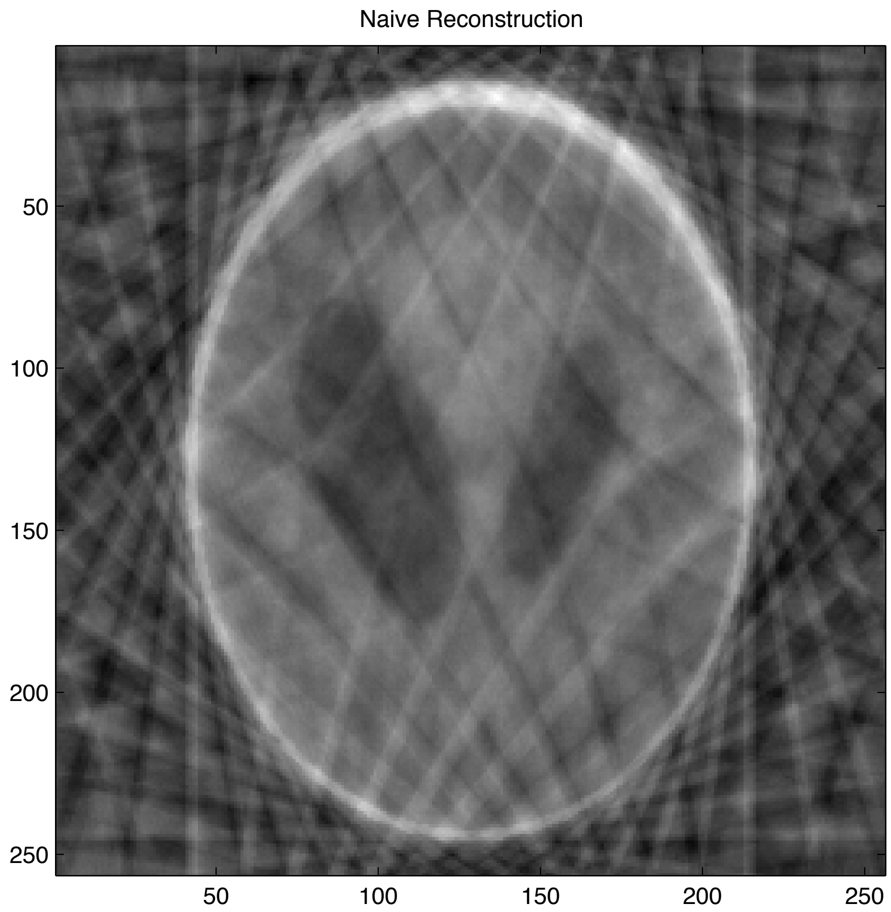


original



g^*

Undersample Nyquist by 25 or 50 at high frequencies!



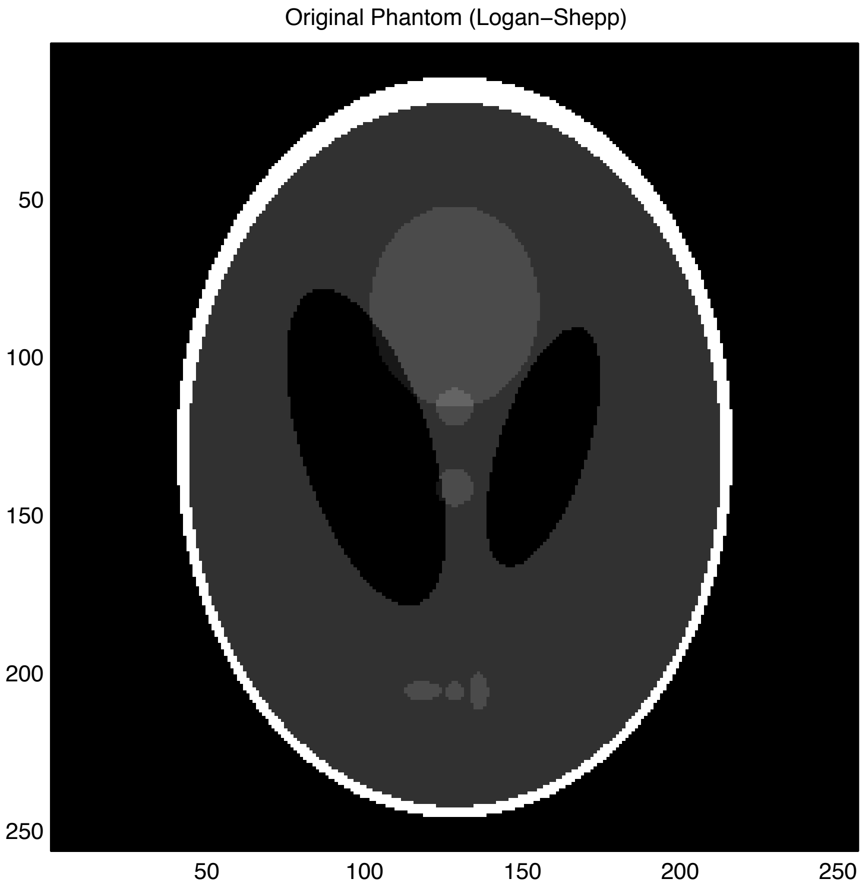
$$\|g\|_{TV} = \sum_{t_1,t_2} |Dg(t_1,t_2)|$$

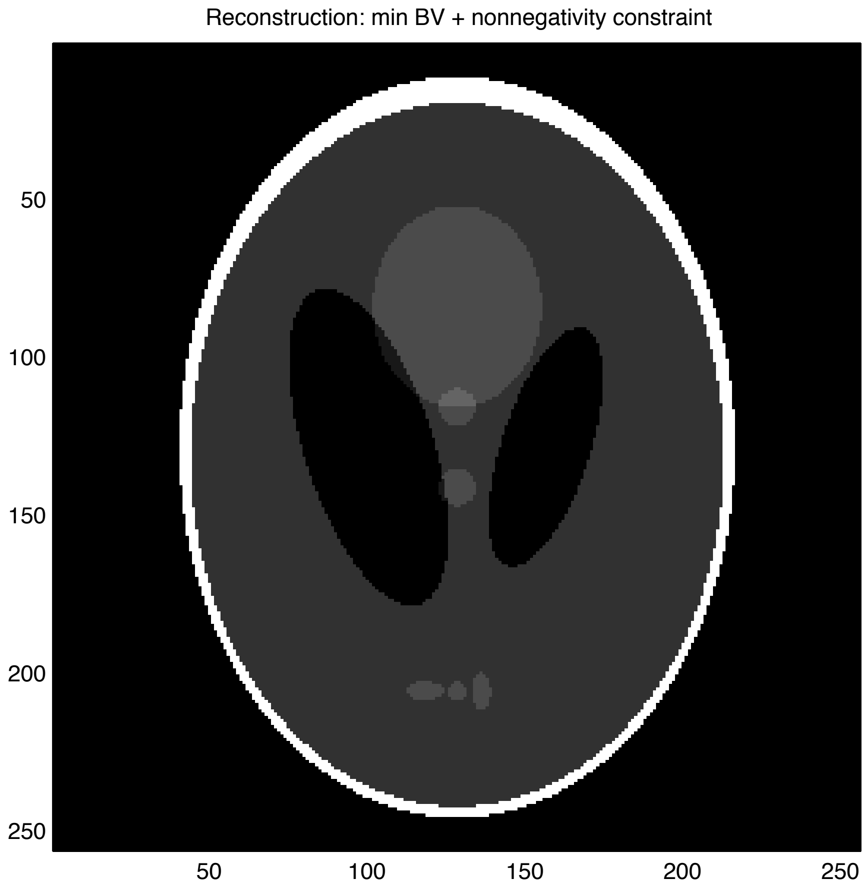
$$Dg(t_1,t_2) = \begin{pmatrix} g(t_1+1,t_2) - g(t_1,t_2) \\ g(t_1,t_2+1) - g(t_1,t_2) \end{pmatrix}$$

Total Variation Reconstruction

Reconstruct g^* with

$$\min_{g} \|g\|_{TV} \quad \text{s.t.} \quad \hat{g}(\omega) = \hat{f}(\omega), \;\; \omega \in \Omega$$





original

$g^* =$ original — perfect reconstruction!

Agenda: Reconstruction of signals from undersampled data

- A New Nonlinear Sampling theory
- Compressive sampling
- Robustness
- Numerical evidence
- Implications and opportunities

Areas of Preoccupation

Wish to recover a digital signal (vector) $x(t) \in \mathbf{R}^N$ from undersampled data:

$$y_k = \langle x, \phi_k \rangle, \quad k = 1, \dots, K, \qquad y = \Phi x$$

Example

- x: 1D signal
- measure Fourier coefficients of x

$$y_k = \hat{x}(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t) e^{-i 2\pi \omega_k t / N}$$

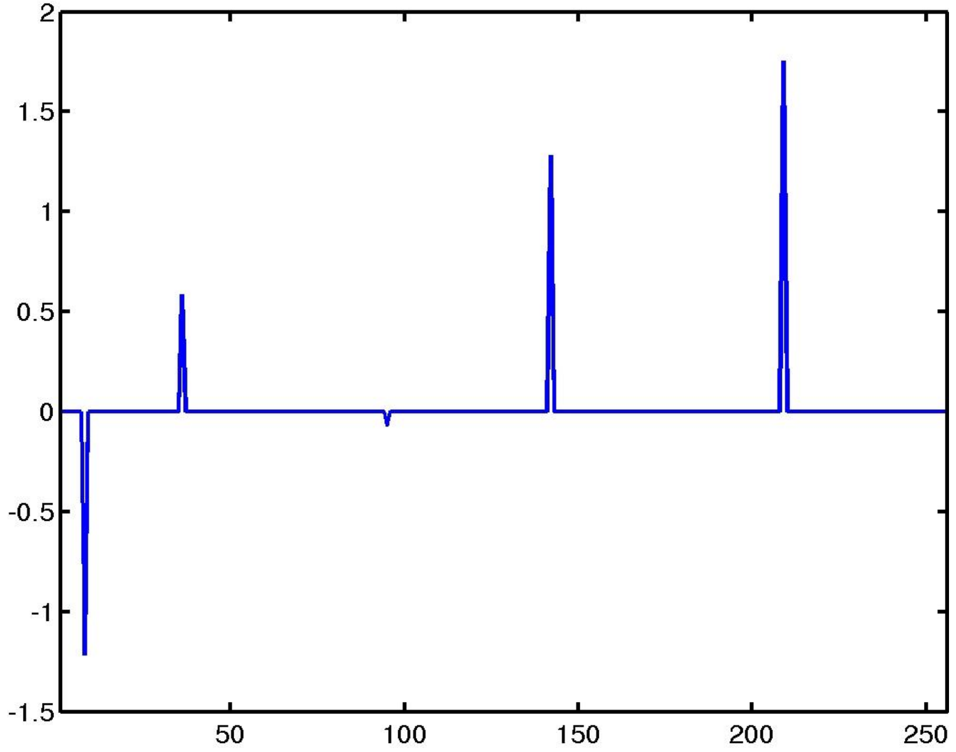
- vastly undersampled data

$$K \ll N$$

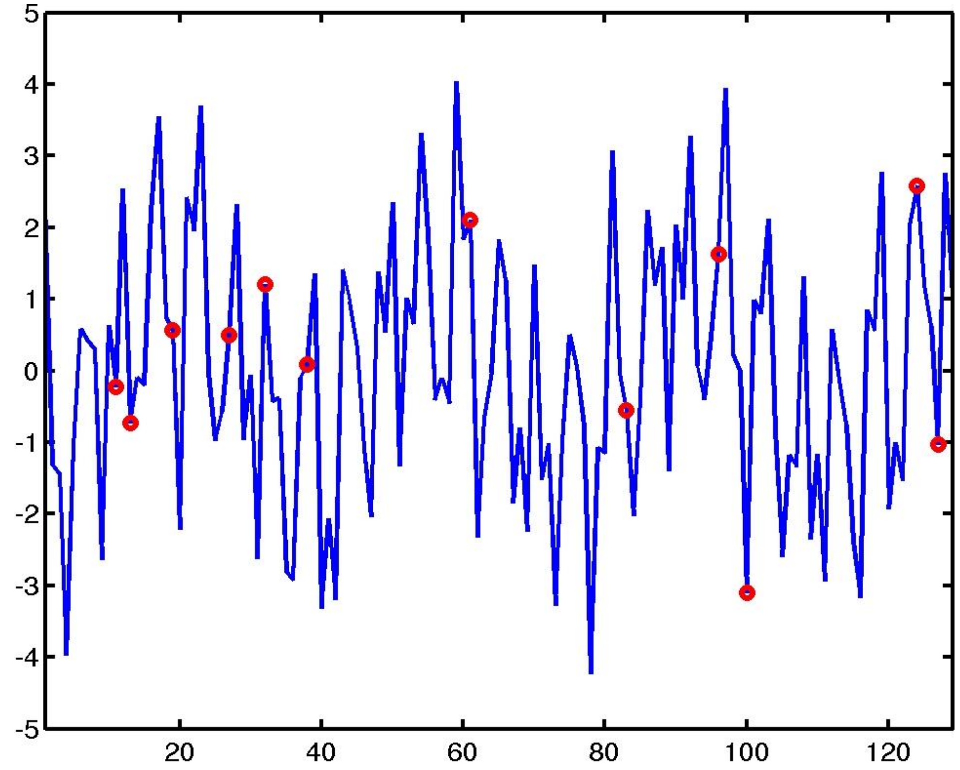
Nyquist says this is impossible: the number of Fourier samples we need to acquire must match the desired resolution N.

Sparse Spike Train

Sparse sequence of $|T|$ spikes



Observe $|\Omega|$ Fourier coefficients

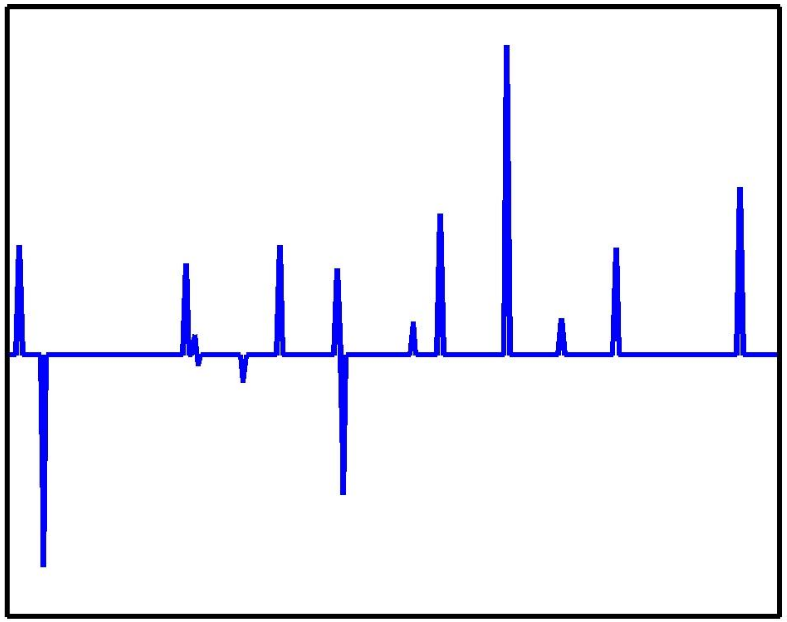


ℓ_1 Reconstruction

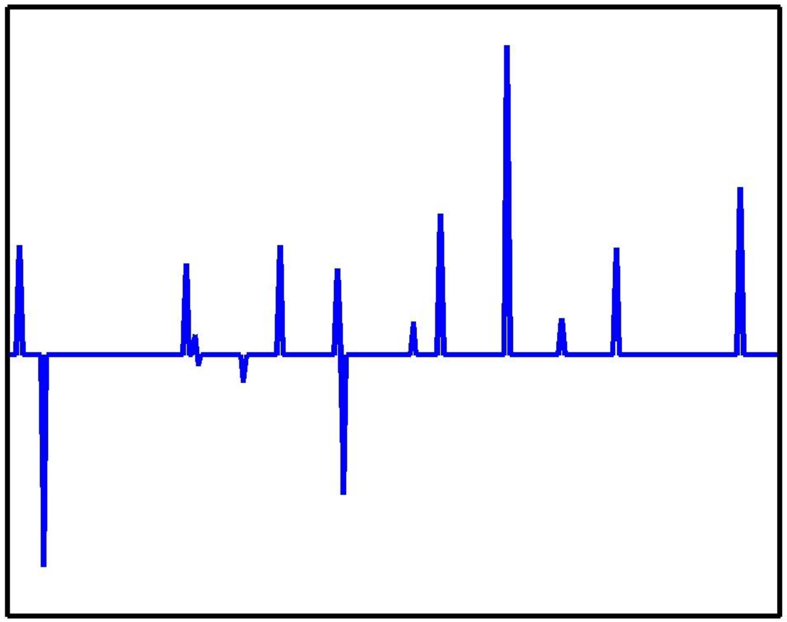
Reconstruct by solving

$$\min_s \|s\|_{\ell_1} := \sum_t |s(t)| \quad \text{s.t.} \quad \Phi s = y$$

$$\hat{s}(\omega_k) = y_k$$



original



recovered from 30 Fourier samples

Min ℓ_1: Equivalent Formulation

$$\min_{s} \|s\|_{\ell_1} := \sum_t |s(t)| \quad \text{subject to} \quad \Phi s = y$$

Reformulation as a linear program (LP)

$$\min \sum_t u(t) \quad \text{s.t.} \quad -u(t) \leq s(t) \leq u(t), \; \Phi s = y$$

with variables $u, s \in \mathbf{R}^N$.

A Nonlinear Sampling Result

- x supported on set of size B
- K frequencies selected at random

$$K \gtrsim B \log N.$$

Minimizing ℓ_1 reconstructs exactly with overwhelming probability.

- We do not need more samples
- We can't do with fewer samples
- In practice, works for *most* with about $|\Omega| = 2|T|$ samples.
- Hard analysis
- Many extensions: higher dimensions, piecewise constant signals, etc. E.g. for 2D TV, the gradient is sparse

$$K \gtrsim \#\text{Jumps} \cdot \log N.$$

Nonlinear Sampling Theorem

- Switch roles of time and frequency:
  - $\hat{x}$ supported on set Ω in freq domain
  - Sample on set T in time domain
- Nonlinear sampling theorem:
  - Ω is an *arbitrary* set of size B
  - We can reconstruct from $\sim B \log N$ *randomly* placed samples
  - Nonlinear reconstruction by convex programming
- Partial conclusions
  - Nyquist is irrelevant
  - Nonlinear sampling theory based on structural content of the signal

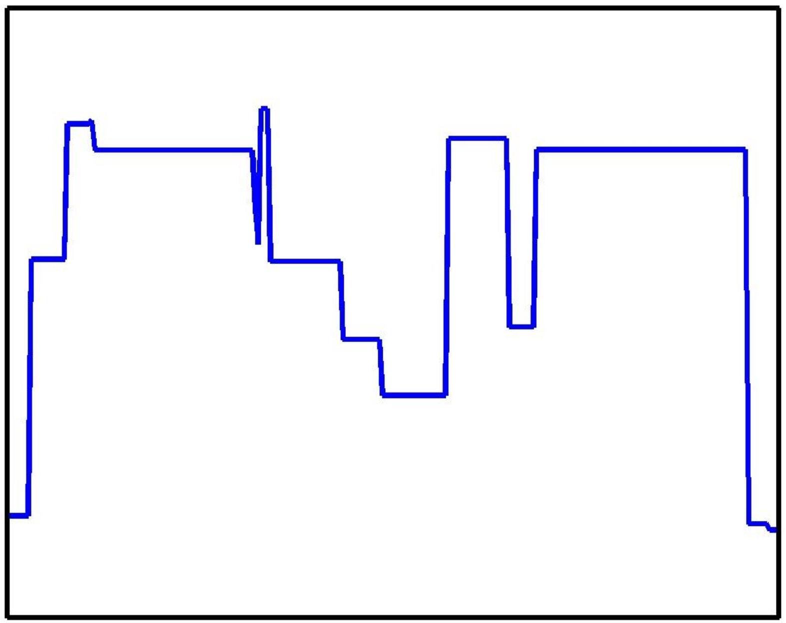
Extensions

- B number of jump discontinuities (TV reconstruction)
- B number of 2D, 3D spikes.
- B number of 2D jump discontinuities (2D TV reconstruction)

E.g. for 2D TV, the gradient is sparse and if

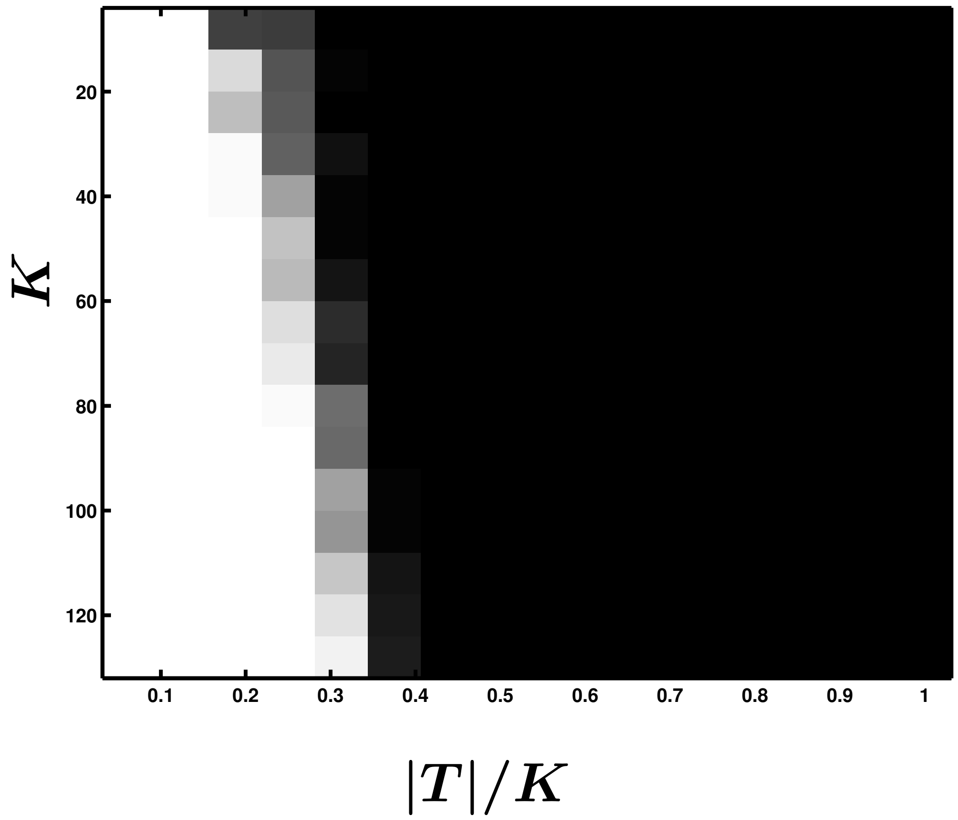
$$K \gtrsim \#B \cdot \log N,$$

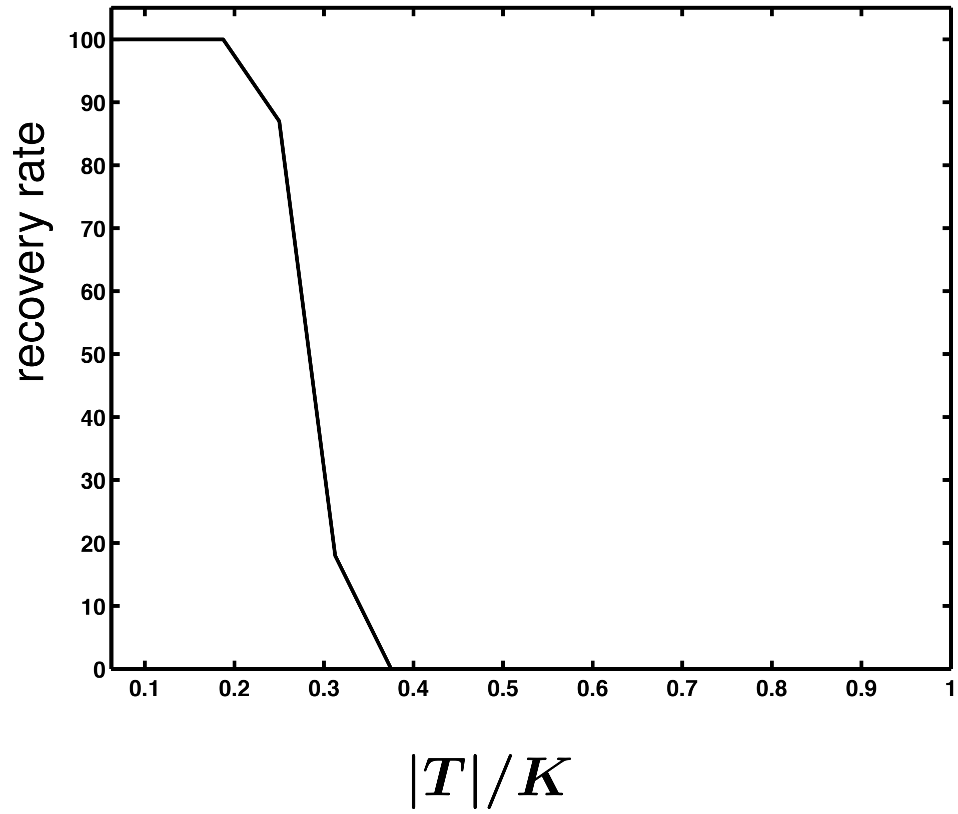
the min-TV is exact with high probability.



Numerical Results

- Signal length $N = 1024$
- Randomly place $|T|$ spikes, observe K random frequencies
- Measure $\%$ recovered perfectly
- white = always recovered, black = never recovered





Previous Work

- ℓ_1 reconstruction in widespread use
  - Santosa and Symes (1986), and others in Geophysics (Claerbout)
  - Donoho and Stark
- Sparse decompostions via Basis Pursuit
  - Chen, Donoho, Saunders (1996)
  - Donoho, Huo, Elad, Gribonval, Nielsen, Fuchs, Tropp (2001-2005)
- Novel sampling theorems
  - Bresler and Feng (2002)
  - Vetterli and others (2002-2004)
- Fast algorithms
  - Gilbert, Strauss, et al. (2002-2005)

An Interesting Problem

- Is it possible to reconstruct signals of scientific interest from a limited number of measurements?
- Which measurements should we take?
- How should we reconstruct?

Signal Recovery from Undersampled Data?

- Think of $x \in \mathbf{R}^N$ as the coefficients of a signal f in an orthogonal basis Ψ

$$f(t) = \sum_{m=1}^{N} x_m \psi_m(t)$$

- Undersampled measurements

$$y_k = \langle f, \phi'_k \rangle, \quad y = \Phi' f,$$

or

$$y_k = \langle x, \phi_k \rangle, \quad \Phi = \Phi' \Psi^*$$

- Recover via LP

$$\min \|s\|_{\ell_1} \qquad \Phi s = y$$

Compressible Signals

- In real life, signals of interest may not be sparse but compressible
- Compressible signal: f is well-approximated by a sparse signal.
- f_B: best B-term expansion

$$f_B(t) = \sum_{B \text{ largest coeff's}} x_m \psi_m(t), \quad f_B = \Phi x_B$$

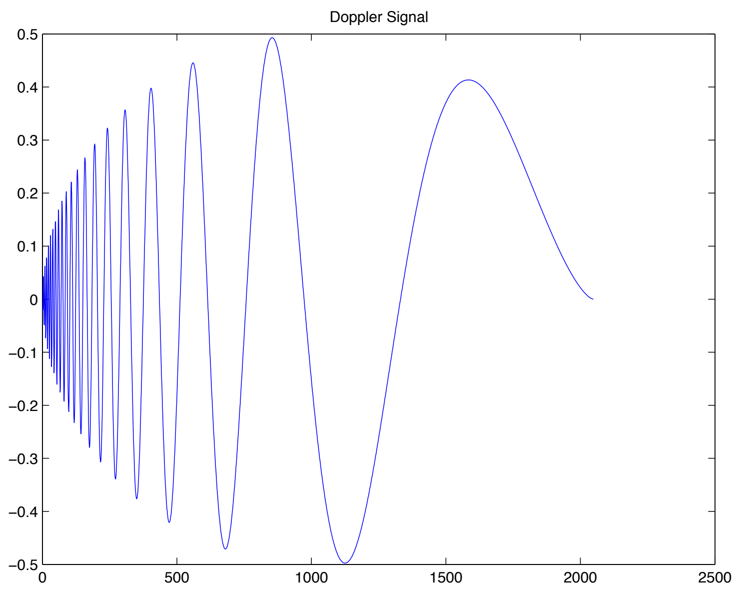
  f_B is a B-sparse signal

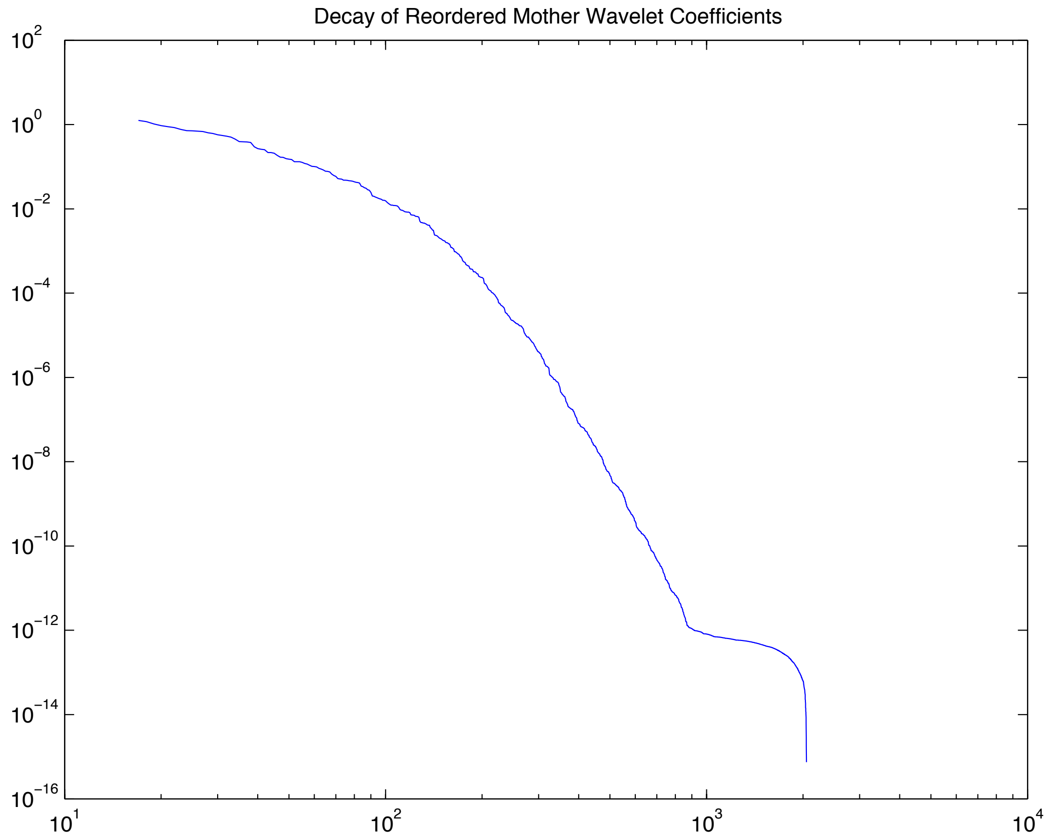
- Many signals are well approximated by a B-sparse signal

$$\|f - f_B\| = \|x - x_B\|$$

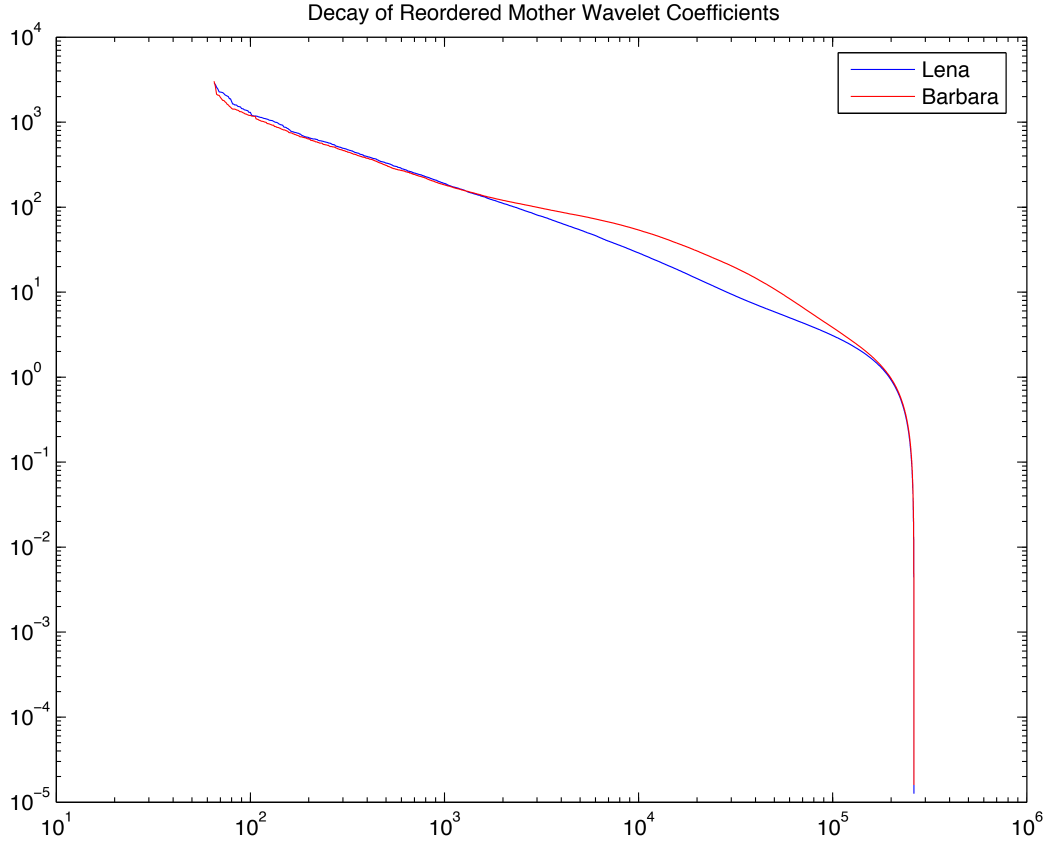
- This is what makes transform coders work (sparse coding)

Compressible Signals I: Wavelets in 1D

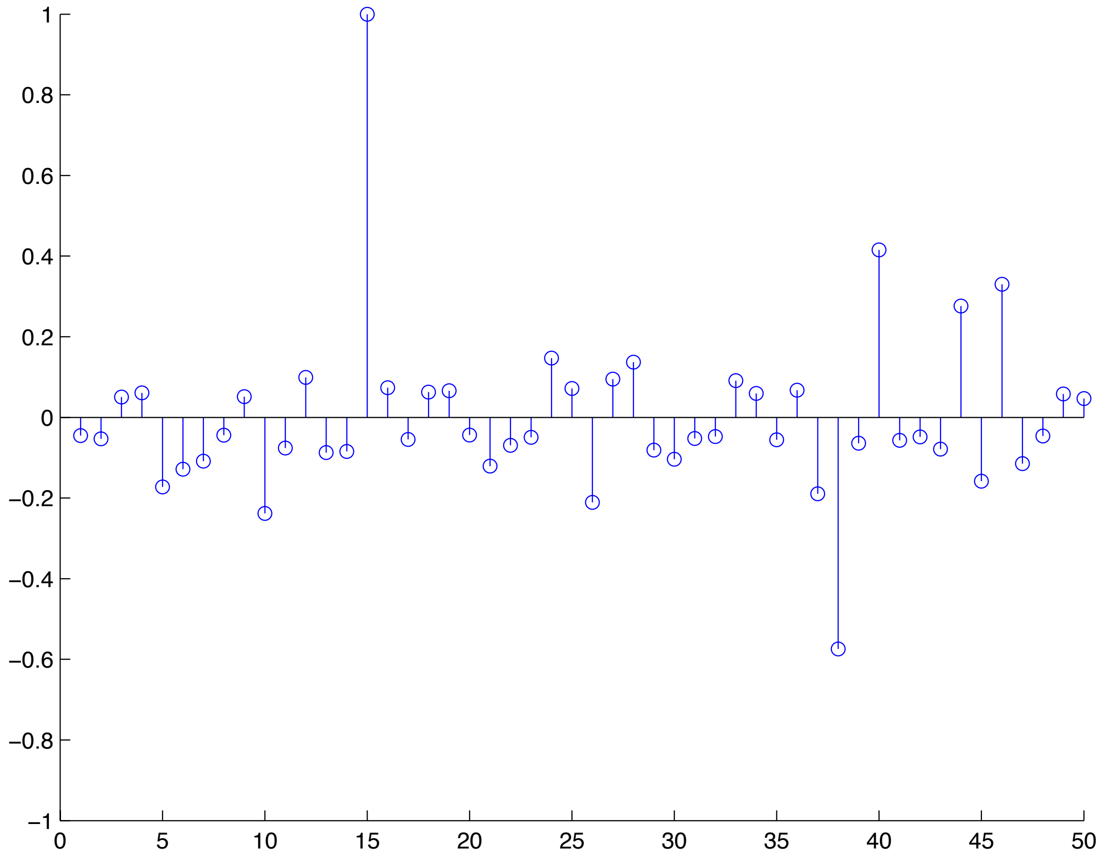




Compressible Signals II: Wavelets in 2D



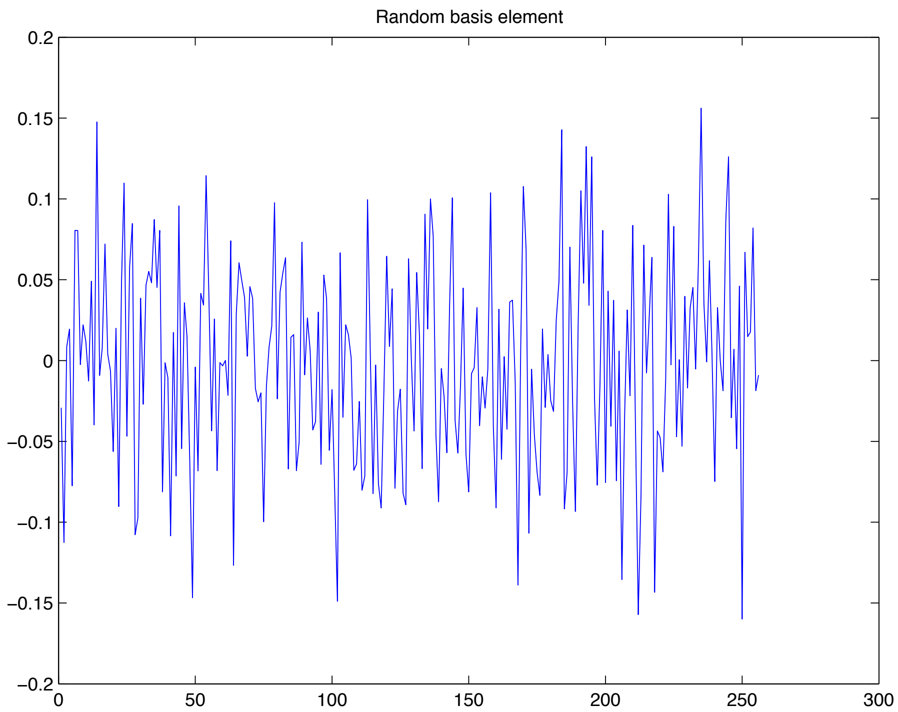
Sensing?

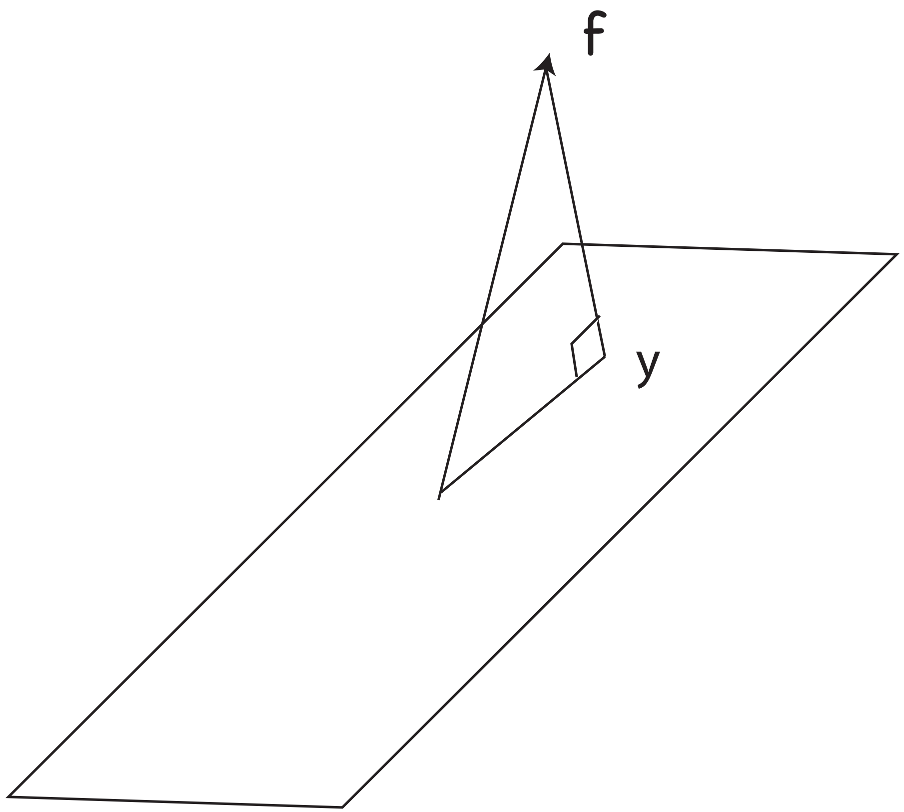


$\boldsymbol{x}$

Random Projections

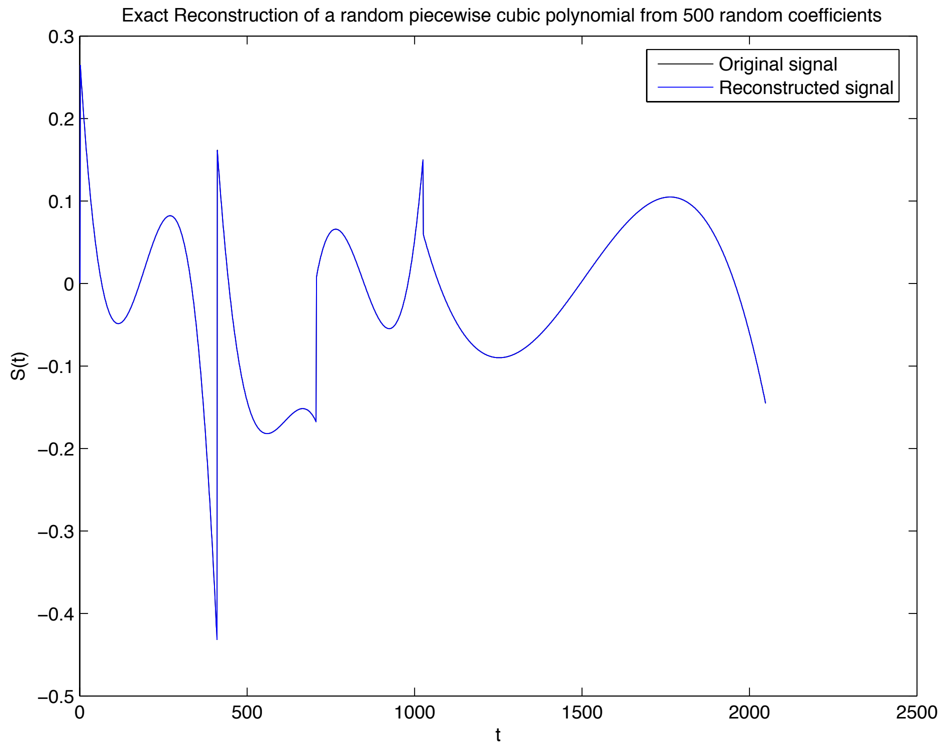
$$y_k = \langle f, X_k \rangle, \quad X_k(t) \ i.i.d. \ N(0,1),$$





Reconstruction of Piecewise Polynomials, I

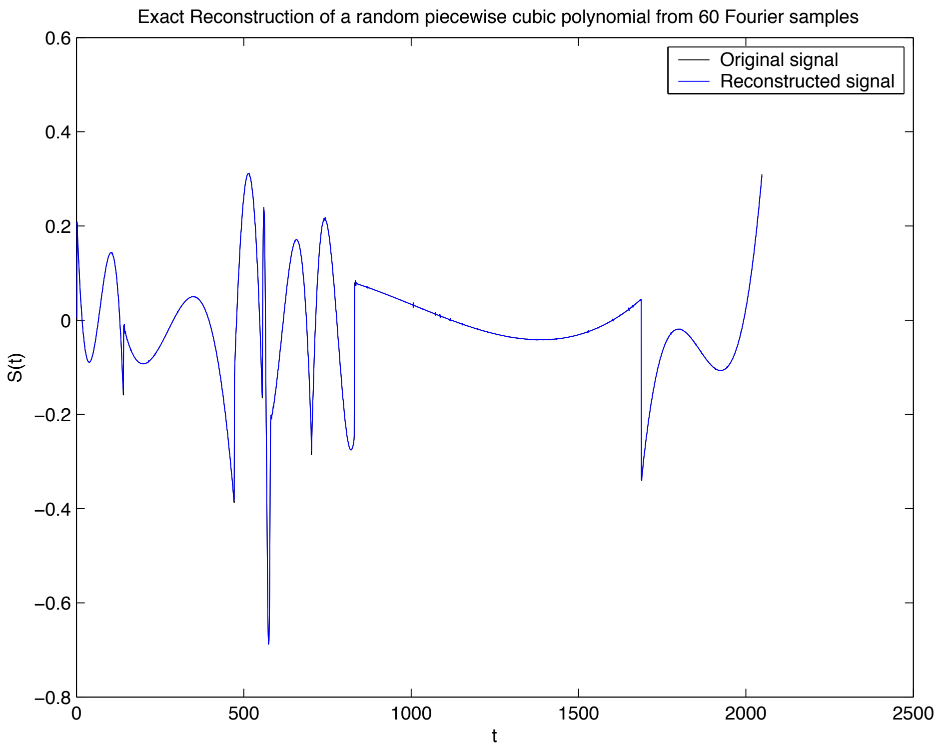
- Randomly select a few jump discontinuities
- Randomly select cubic polynomial in between jumps
- Observe about 500 random coefficients
- Minimize ℓ_1 norm of wavelet coefficients (170 nonzeros)



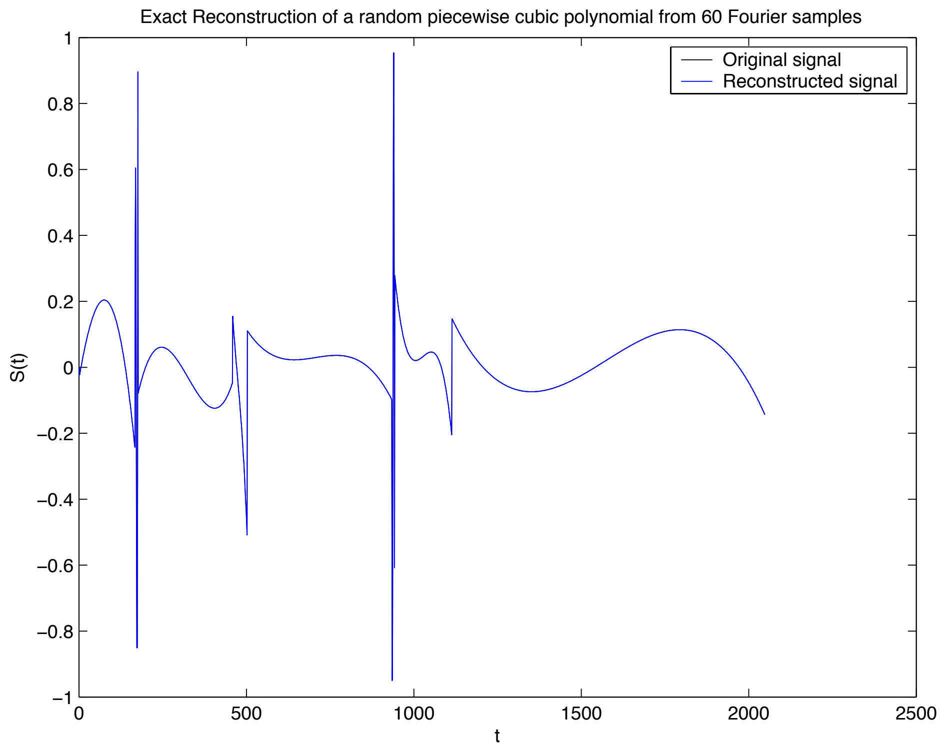
Reconstructed signal

Reconstruction of Piecewise Polynomials,II

- Randomly select 8 jump discontinuities
- Randomly select cubic polynomial in between jumps
- Observe about 200 Fourier coefficients at random



Reconstructed signal



Reconstructed signal

Key Idea: Uncertainty Relation

W. Heisenberg, 1901-1976

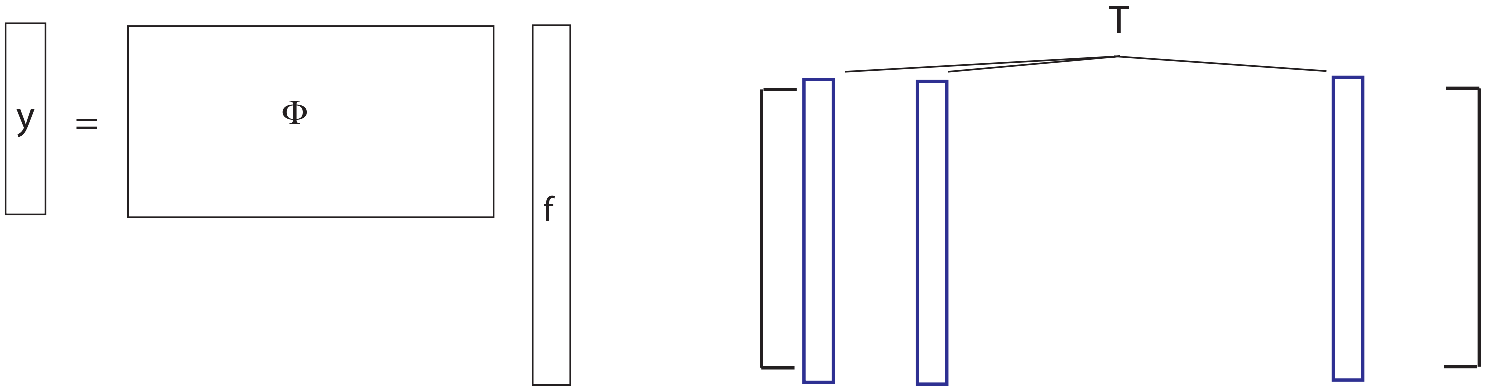
Weyl-Heisenberg Uncertainty Principle

- f 'lives' on an interval of length Δt
- $\hat{f}$ 'lives' on an interval of length $\Delta \omega$

$$\Delta t \cdot \Delta \omega \geq 1$$

Restricted Isometries

- $\boldsymbol{\Phi_T}$ columns of $\boldsymbol{\Phi}$ corresponding to $\boldsymbol{T}$, $\boldsymbol{\Phi_T} \in \mathbf{R}^{\boldsymbol{K} \times |\boldsymbol{T}|}$.



- Restricted isometry constants $\boldsymbol{\delta_S}$

$$(\mathbf{1} - \boldsymbol{\delta_S})\,\mathrm{Id} \le \boldsymbol{\Phi_T^* \Phi_T} \le (\mathbf{1} + \boldsymbol{\delta_S})\,\mathrm{Id}, \qquad \forall \boldsymbol{T},\ |\boldsymbol{T}| \le \boldsymbol{S}.$$

- Sparse subsets of column vectors are approximately orthonormal.

- $\boldsymbol{\Phi}$ obeys a *uniform uncertainty principle* with oversampling factor $\boldsymbol{\lambda_\Phi}$ if $\boldsymbol{\delta_S} \le \mathbf{1/2}$ for $\boldsymbol{S} \le \boldsymbol{K/\lambda_\Phi}$.

- Uniform because must hold for *all* $\boldsymbol{T}$'s.

Why Do We Call This an Uncertainty Principle?

- $\boldsymbol{\Phi = F_\Omega}$, rows of the DFT isometry (corresponding to $\boldsymbol{\Omega}$)
- $\boldsymbol{F_{\Omega T}}$, columns of $\boldsymbol{F_\Omega}$ (corresponding to $\boldsymbol{T}$)
- UUP

$$\boldsymbol{(1-\delta_S)\,\frac{|\Omega|}{N}\cdot\|f_T\|^2 \le \|F_{\Omega T} f_T\|^2 \le (1+\delta_S)\,\frac{|\Omega|}{N}\cdot\|f_T\|^2}$$

- Implications
  - $\boldsymbol{f}$ supported on $\boldsymbol{T}$, $|\boldsymbol{T}| \le \boldsymbol{S}$
  - If UUP holds, then

$$\boldsymbol{(1-\delta_S)\,\frac{|\Omega|}{N} \le \|\hat{f}\cdot 1_\Omega\|^2/\|\hat{f}\|^2 \le (1+\delta_S)\,\frac{|\Omega|}{N}}$$

Signal Recovery from Undersampled Data

$$\min \|s\|_{\ell_1} \qquad \Phi s = \Phi x.$$

- Φ obeys a uniform uncertainty principle with oversampling factor λ_Φ
- Take $B \cdot \lambda_\Phi$ measurements
- For all $x \in \mathbf{R}^N$, (nearly)

$$\|x^\sharp - x\| \leq 8 \, \|x - x_B\|$$

- Useful if oversampling factor is small

Examples: Random Projections

- Gaussian ensemble: entries of Φ are i.i.d. $N(0, 1/K)$

$$\lambda_\Phi \lesssim \log[N/K]$$

- Binary ensemble: entries of Φ are i.i.d. $\pm 1/\sqrt{K}$

$$\lambda_\Phi \lesssim \log[N/K]$$

- Fourier ensemble: select K random Fourier coefficients (see also Rudelson and Vershynin, 2005)

$$\lambda_\Phi \lesssim \log^4 N$$

All with overwhelming probability.

Signal Recovery from Gausssian Measurements

$$\min \|s\|_{\ell_1} \qquad \Phi s = \Phi x.$$

- The measurement vectors are i.i.d. white noise
- Take about $B \cdot \log(N/B)$ measurements
- For all $x \in \mathbf{R}^N$, (nearly)

$$\|x^\sharp - x\| \leq 8 \, \|x - x_B\|$$

- See also Donoho, 2004.

Big Surprise!

Want to sense an object for accurate reconstruction

- *Strategy 1:* Oracle tells exactly which coefficients are large (collect all N wavelet coefficients, sort them and select the largest)
- *Strategy 2:* Collect $B\log[N/B]$ random coefficients and reconstruct by LP.

Surprising claim

- Same performance
- Performance is achieved by solving an LP.

See also Donoho, 2004.

Optimality

- Can you do with fewer than $B\log[N/B]$ for same accuracy?
- Simple answer: NO
- Connected with theory of Gelfand widths
- Connected with information theory (rate-distortion curve of compressible signals)

Compressive Sampling (CS)

CS suggests "the possibility of compressed data acquisition protocols which perform as if it were possible to directly acquire just the important information about the image of interest."

Dennis Healy, DARPA

Incoherence

- $f = \sum_m x_m \psi_m(t)$
- K randomly selected coefficients in basis another orthobasis Φ

$$y_k = \langle f, \phi_k \rangle \quad k \in \Omega \quad y = \Phi_\Omega f.$$

- Recover via

$$\min \|s\|_{\ell_1} \qquad \Phi_\Omega \Psi^* s = y$$

- Oversampling factor is about

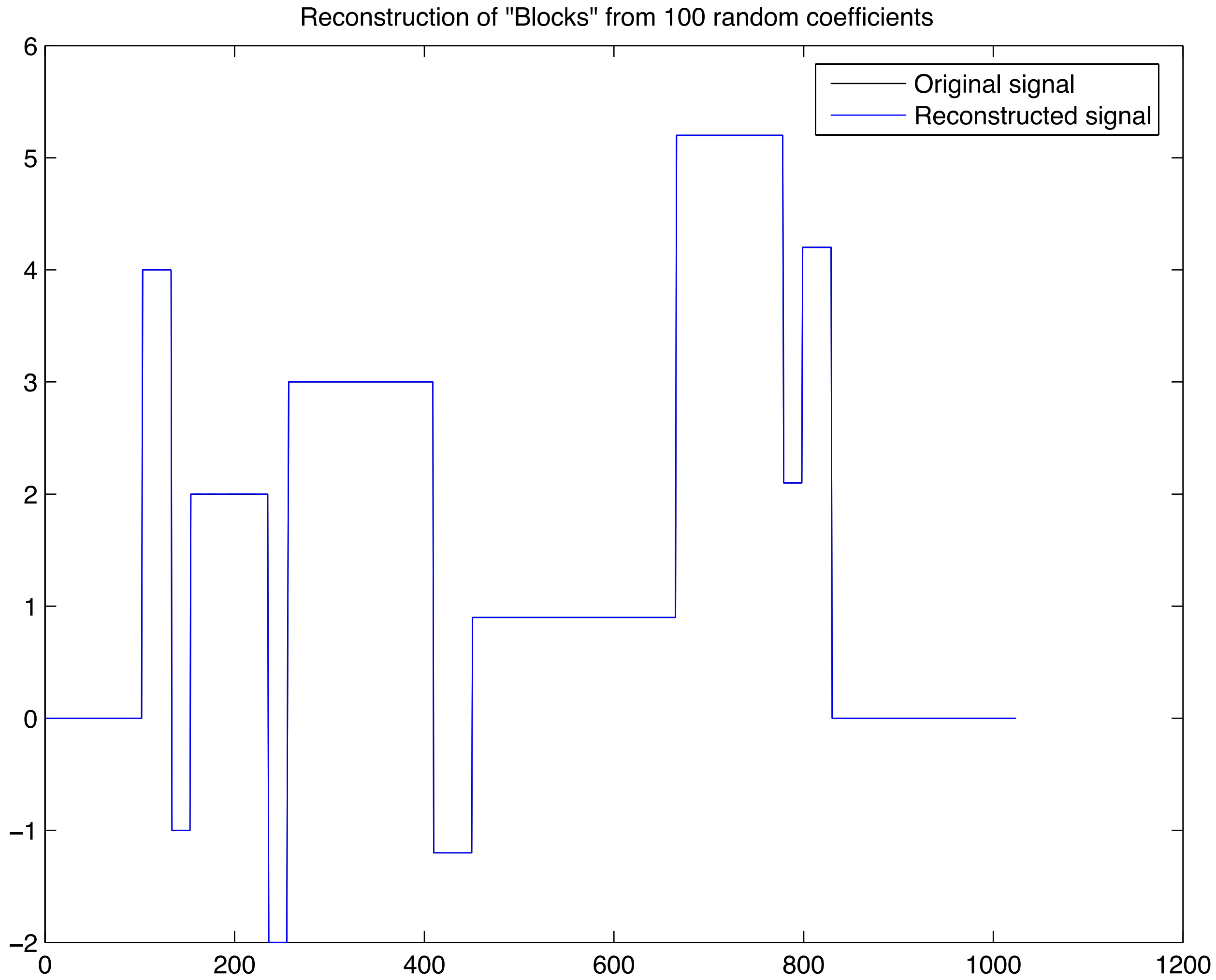
$$\mu^2 (\log N)^4$$

I.e. we need $K \gtrsim B \cdot \mu^2 (\log N)^4$ to recover B largest terms.

- Mutual coherence

$$\mu = \sqrt{N} \, \max |\langle \phi_k, \psi_m \rangle|, \quad 1 \leq \mu \leq \sqrt{N}$$

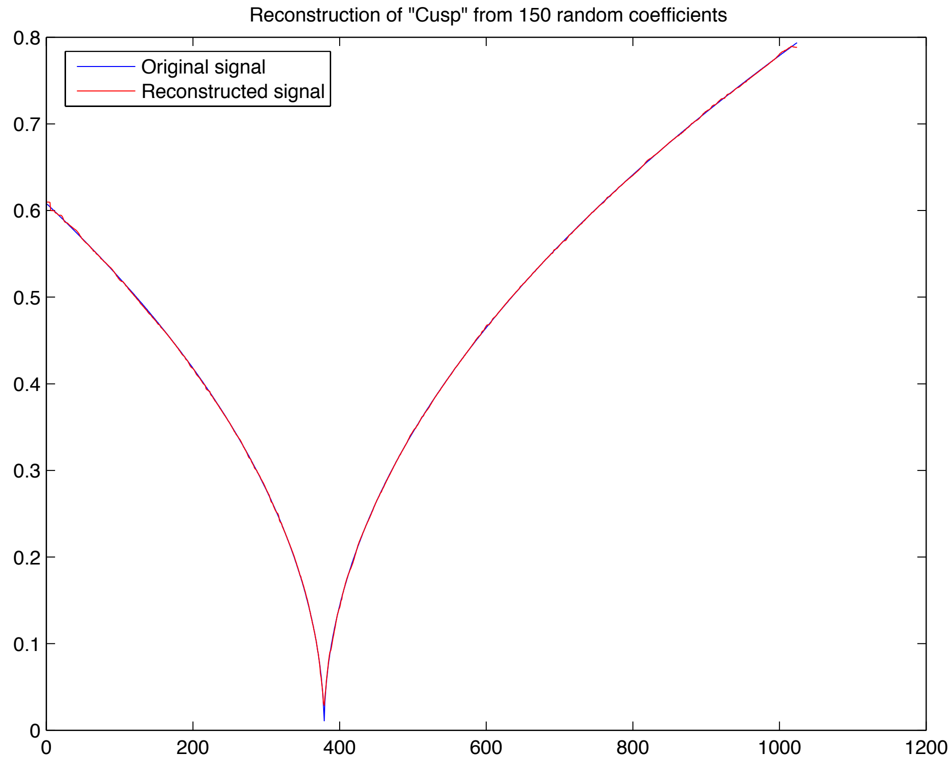
Reconstruction from 100 Random Coefficients



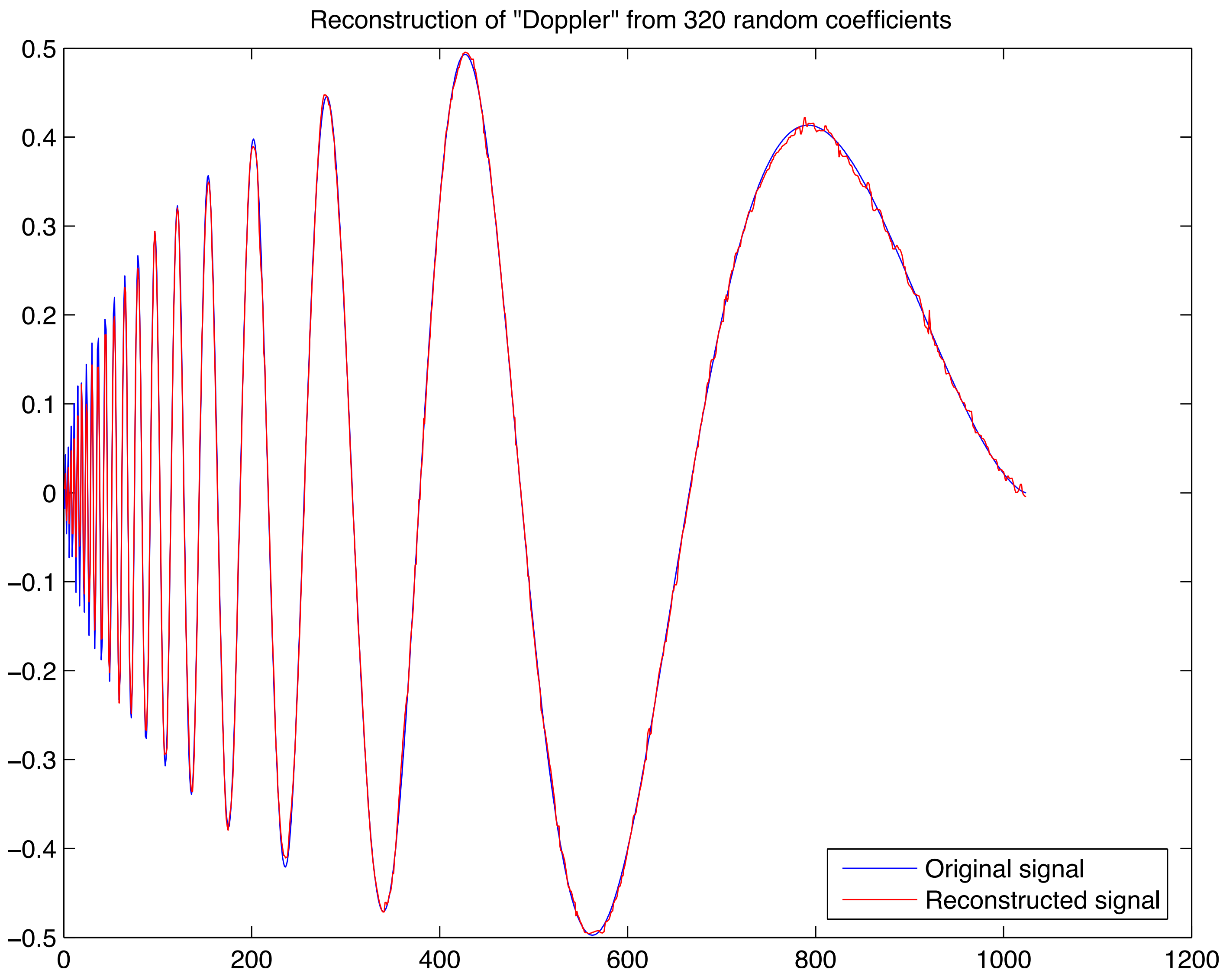
'Compressed' signal

Reconstruction from Random Coefficients

Minimize TV subject to random coefficients + ℓ_1-norm of wavelet coefficients.



Reconstruction from Random Coefficients



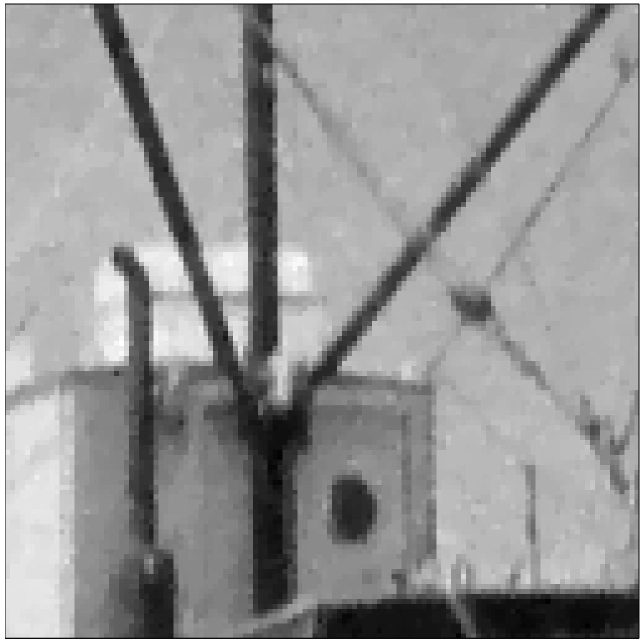
original, 65k pixels

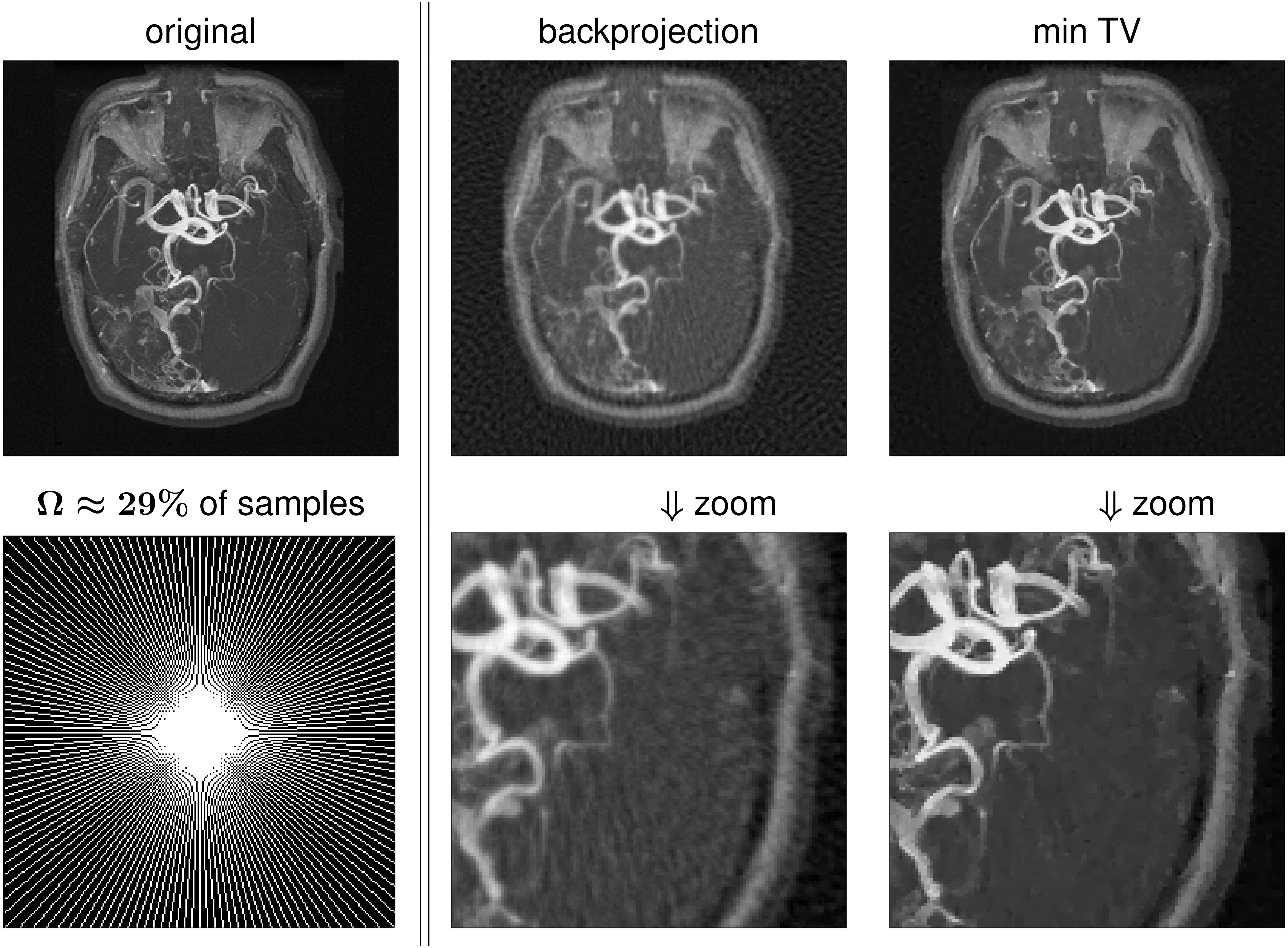
wavelet 7207-term approx

recovery from 20k proj

$\Downarrow$ zoom

$\Downarrow$ zoom




original
backprojection
min TV
$\Omega \approx 29\%$ of samples
⇓ zoom
⇓ zoom

Stable Recovery?

- In real applications, data are corrupted
- Better model: $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$, where $\boldsymbol{e}$ may be stochastic, deterministic.
- Recall most of the singular values of $\boldsymbol{\Phi}$ are zero
- Hopeless?

Stable Recovery from Undersampled Data

$$\min \|s\|_{\ell_1} \qquad \|y - \Phi s\|_{\ell_2} \leq \|e\|_{\ell_2}$$

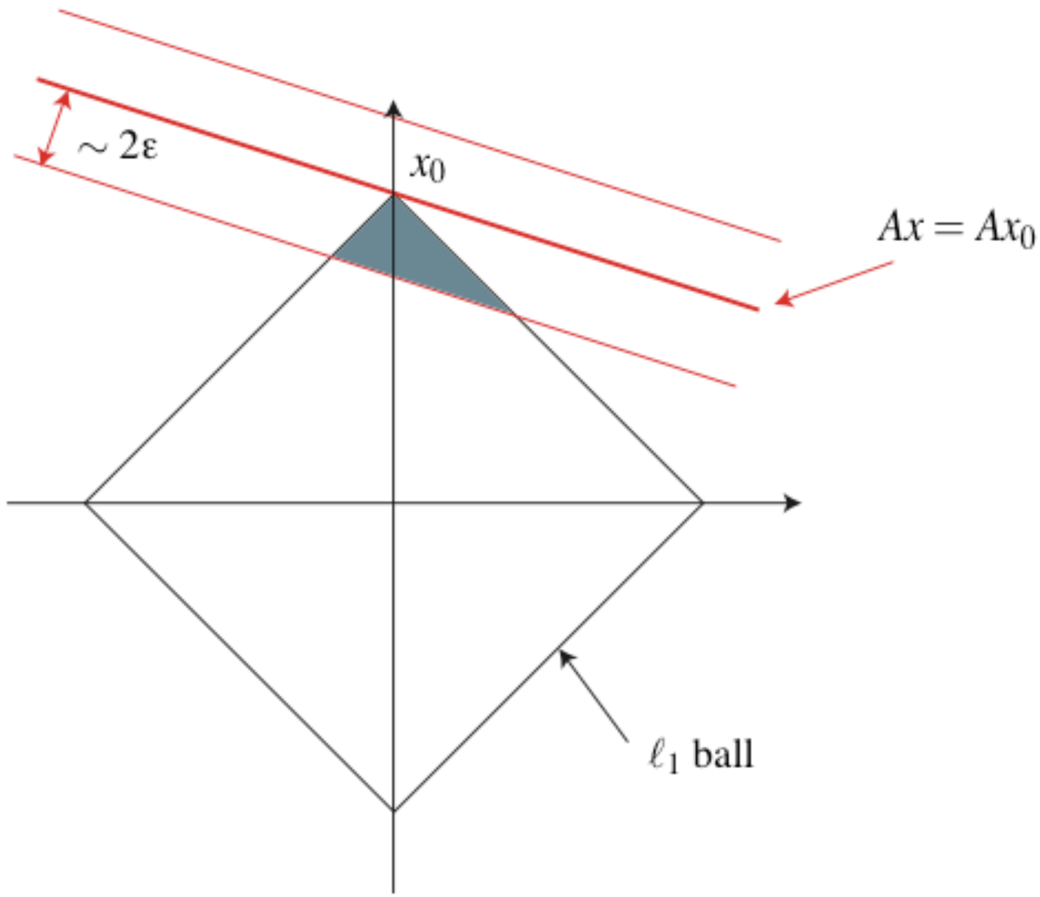
With $K = B \cdot \lambda_\Phi$ (nearly)

$$\|x - x^\sharp\|_{\ell_2} \leq 8 \cdot (\|x - x_B\|_{\ell_2} + \|e\|_{\ell_2})$$

- No blow up!
- Reconstruction within the noise level
- Nicely degrades as noise level increases

Geometric Intuition

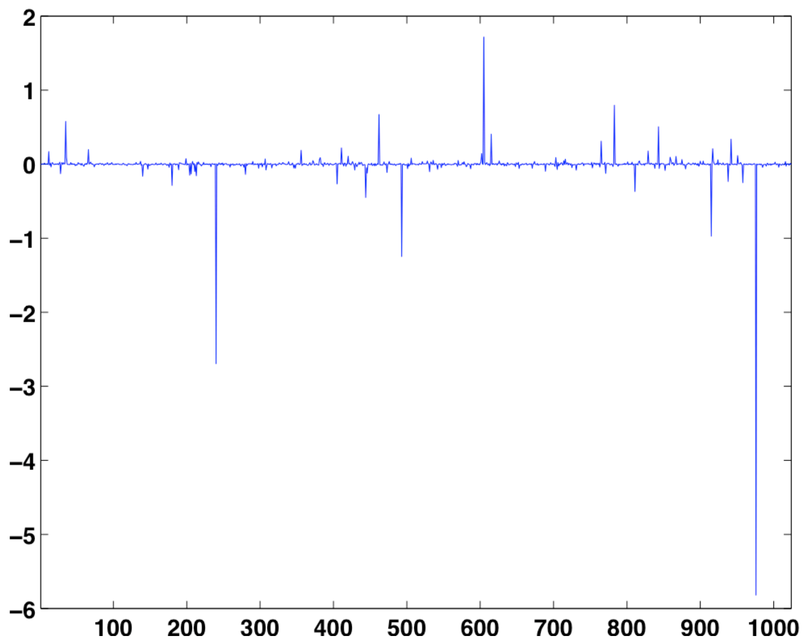
- $x^\sharp$ feasible $\Rightarrow$ $x^\sharp$ inside the diamond
- $x^\sharp$ obeys the constraint $\Rightarrow$ $x^\sharp$ inside the slab (tube)



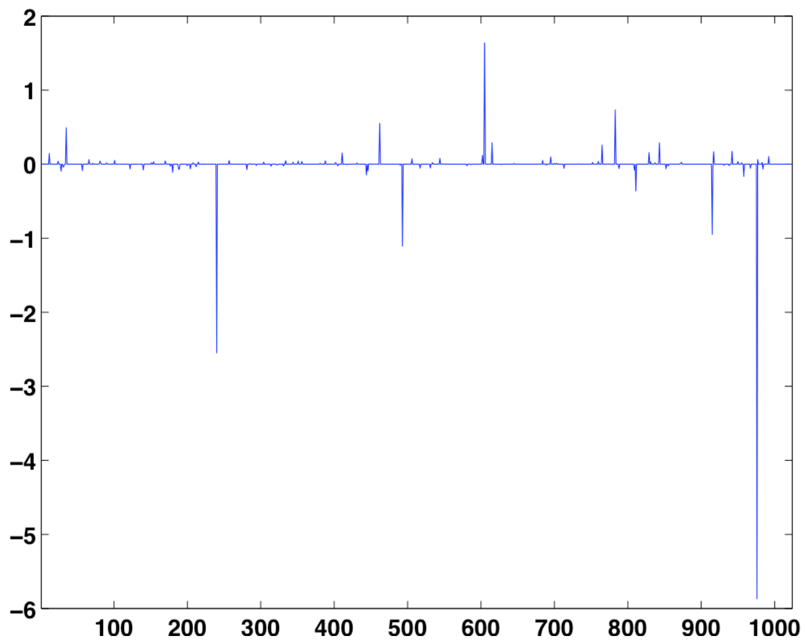
Stable Recovery, I

Recovery of compressible signal $(N = 1024)$: Gaussian white noise of variance σ^2 added to each of the $K = 300$ measurements.

σ	0.01	0.02	0.05	0.1	0.2	0.5
ϵ	0.19	0.37	0.93	1.87	3.74	9.34
$\|x^\sharp - x_0\|_2$	0.69	0.76	1.03	1.36	2.03	3.20



Original



Recovered

Stable Recovery, II

Image recovery from undersampled and contaminated measurements

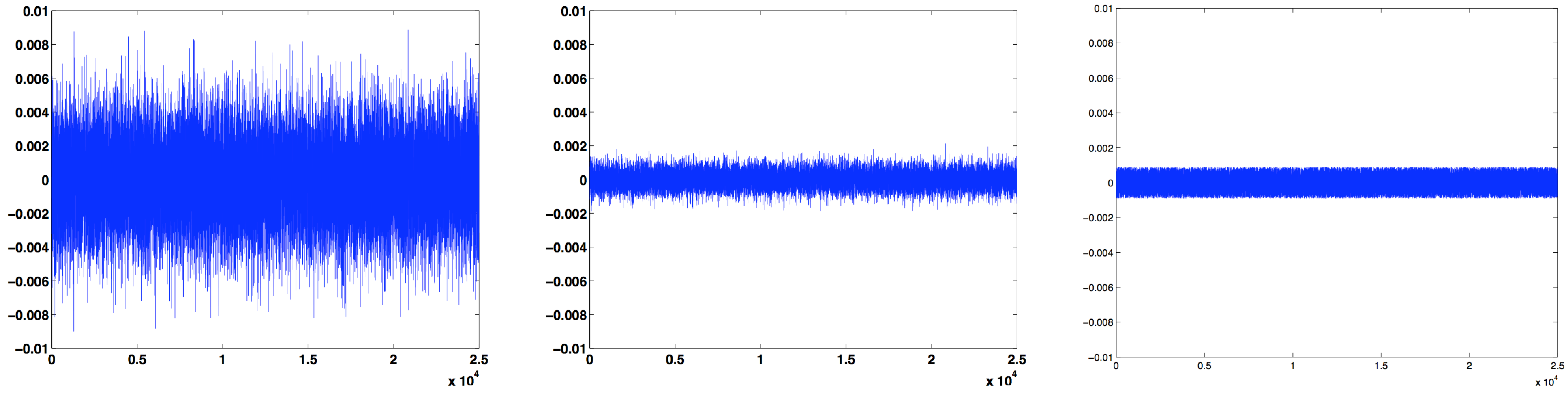
- Gaussian white noise
- Round-off error

	White noise	Round-off
$\|\|e\|\|_{\ell_2}$	0.0789	0.0824
ϵ	0.0798	0.0827
$\|\|\alpha^\sharp - \alpha_0\|\|_{\ell_2}$	0.1303	0.1323
$\|\|\alpha^\sharp_{TV} - \alpha_0\|\|_{\ell_2}$	0.0837	0.0843

(a) (b) (c)

- (b) Recovery via (TV) from 25000 measurements corrupted with Gaussian noise.
- (c) Recovery via (TV) from 25000 measurements corrupted by round-off error.

In both cases, the reconstruction error is less than the sum of the nonlinear approximation and measurement errors.



(a) (b) (c)

- (a) Noiseless measurements $\boldsymbol{Ax_0}$ of the *Boats* image.
- (b) Gaussian measurement error with $\boldsymbol{\sigma = 5 \cdot 10^{-4}}$. The signal-to-noise ratio is $\boldsymbol{\|Ax_0\|_{\ell_2} / \|e\|_{\ell_2} = 4.5}$.
- (c) Round-off error. The signal-to-noise ratio is $\boldsymbol{\|Ax_0\|_{\ell_2} / \|e\|_{\ell_2} = 4.3}$.

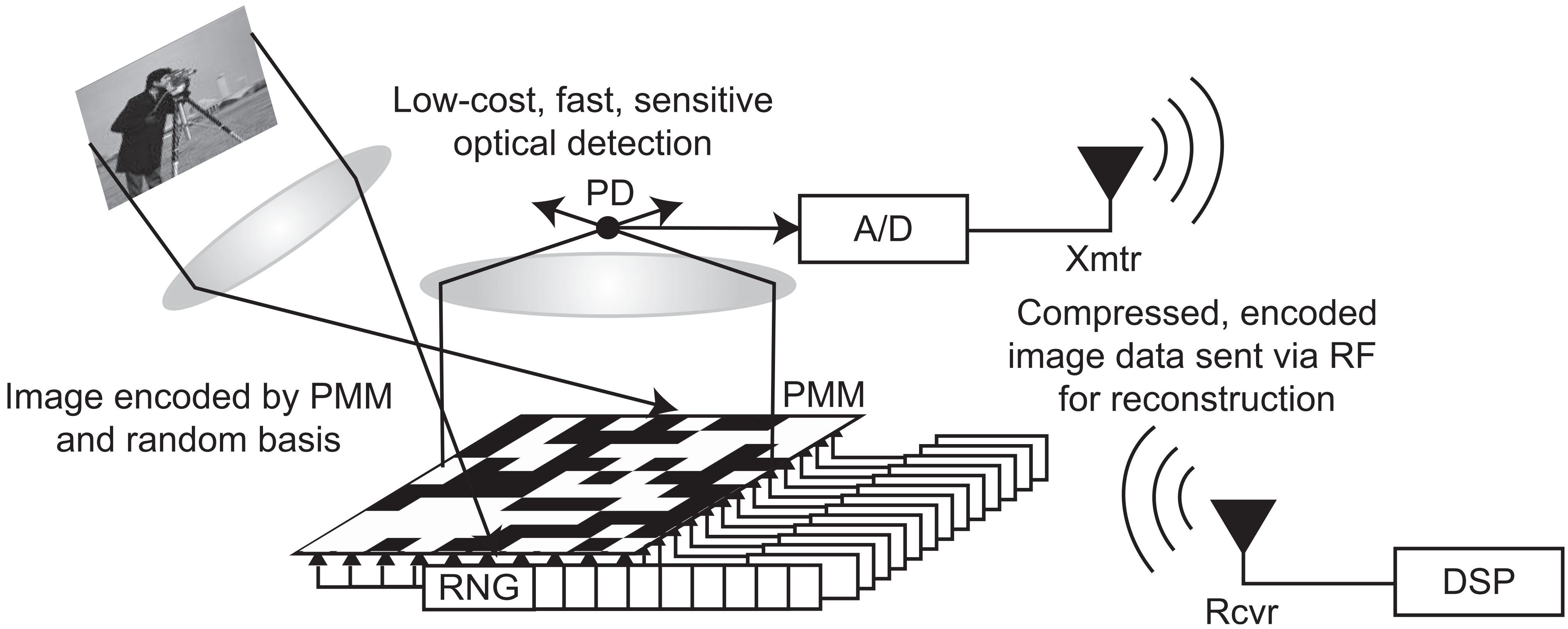
Take Home Message: Power of Random Sensing

- Wish to sense an object
- Take $3B - 4B$ random measurements
- Reconstruct by convex programming
- As accurate as the best B-term expansion in your favorite basis

New Paradigm for Analog to Digital

- Measure K general linear functionals rather than the usual pixels
- Reconstruct with essentially the same resolution as that one would obtain by measuring all the pixels.
- Impact for sensor design: incoherent analog sensors
- Pay-off: far fewer sensors than what is usually considered necessary.

Rice Compressed Sensing Camera



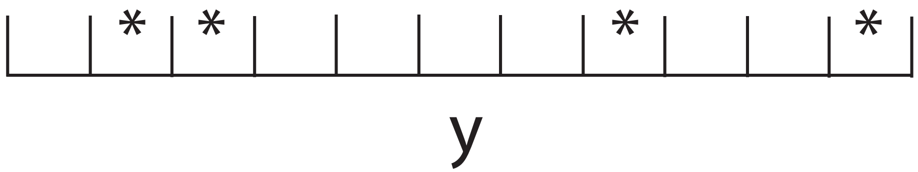
Richard Baraniuk, Kevin Kelly, Yehia Massoud, Don Johnson
dsp.rice.edu/CS

Conclusions

- New nonlinear sampling theory
- Many applications
  - Sparse approximations with overcomplete representations
  - Error-correction
  - Statistical estimation
  - Information theory
  - Super-resolution
- Opportunities
  - Biomedical imagery
  - New A/D devices (Rice University, others)
  - New paradigms for sensor networks (R. Nowak, others)

The Error Correction Problem

- We wish to transmit a "plaintext" $\boldsymbol{x} \in \mathbf{R}^n$ reliably
- Frequently discussed approach: encoding, e.g. generate a "ciphertext" $\boldsymbol{Ax}$, where $\boldsymbol{A} \in \mathbf{R}^{m \times n}$ is a coding matrix
- Assume a fraction of the entries of $\boldsymbol{Ax}$ are corrupted $\longrightarrow \boldsymbol{y}$



  - Corruption is arbitrary
  - We do not know which entries are corrupted
  - We do not know how the corrupted entries are affected
- Is it possible to recover the plaintext exactly from the corrupted ciphertext?

Decoding by Linear Programming

To recover x, solve

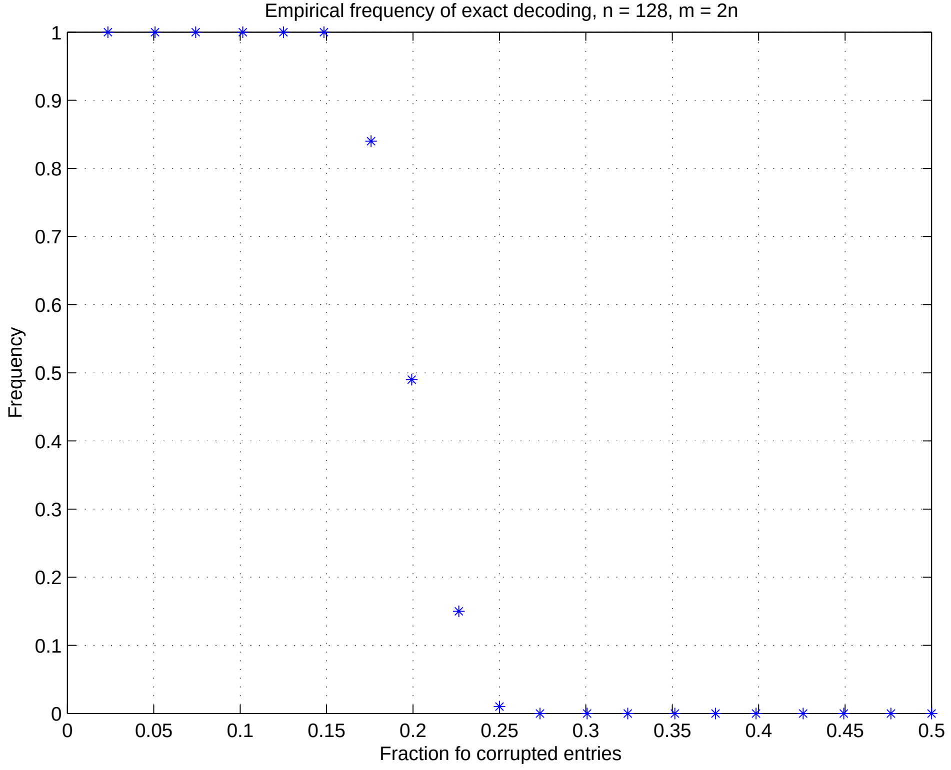
$$\min_{s \in \mathbf{R}^n} \|y - As\|_{\ell_1}.$$

Suppose A is Gaussian, then the input x is the unique solution provided the fraction ρ of corrupted entries is not too large, $\rho \le \rho^*$

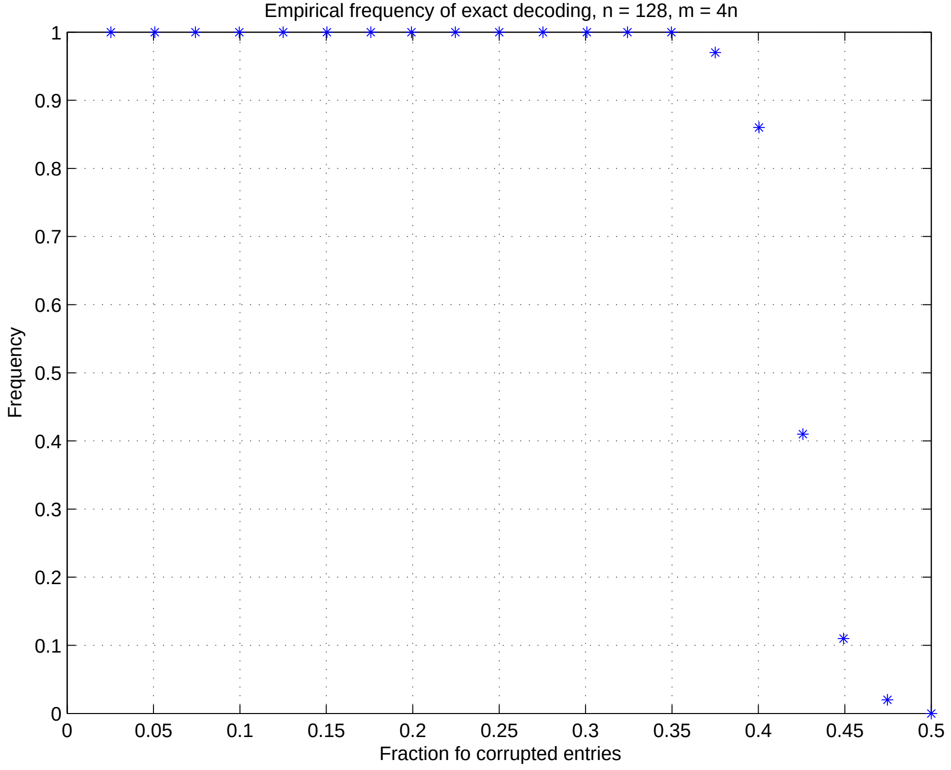
Donoho (2005) showed that for $m = 4n$, $\rho^* = 11\%$!

Practical Performance, I

- A_{ij} i.i.d. $N(0,1)$
- $x \in \mathbf{R}^n$
- Corruption: flip the sign of randomly selected entries



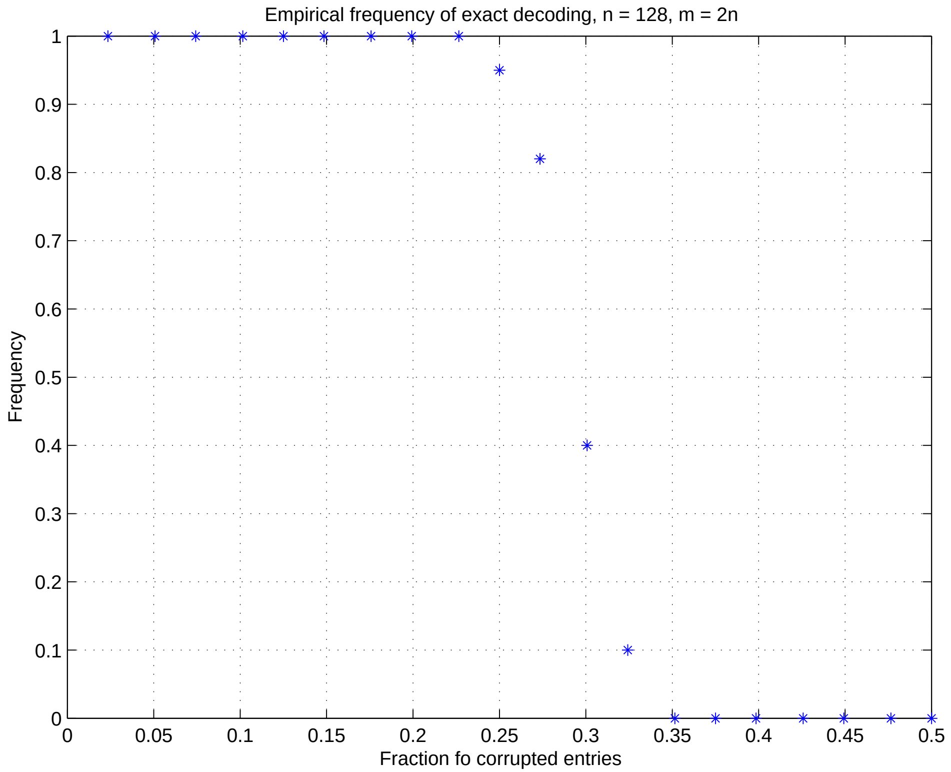
$n = 128,\, m = 2n$



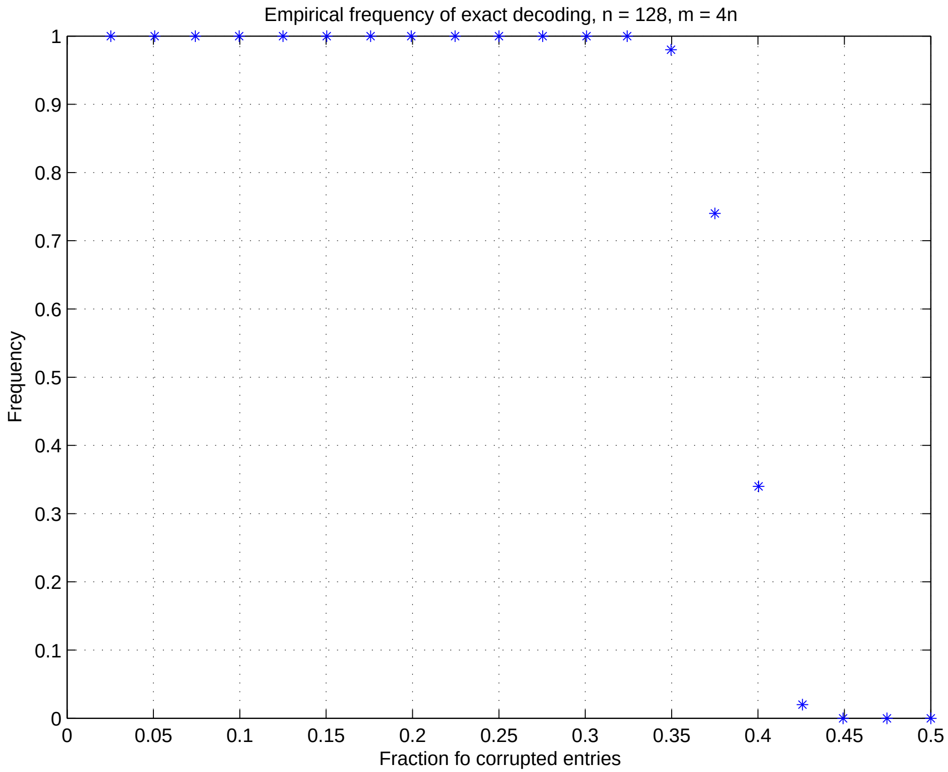
$n = 128,\, m = 4n$

Practical Performance, II

- A_{ij} i.i.d. $\mathbf{P}(A_{ij} = \pm 1) = 1/2$
- $x \in \{0, 1\}^n$
- Corruption: flip the sign of randomly selected entries
- Solve $\min \|y - As\|_{\ell_1}$ subject to $0 \le s \le 1$, and round up.



$n = 128, m = 2n$



$n = 128, m = 4n$